

Chapter 5

Analysing Gaze for RSVP

Abstract In this chapter we analyse 10 distinct RVSP modes drawn from across the whole range of static, moving and multiple entry/exit modes in terms of the visual search, steady gaze, nystagmus and visual pursuit characterisations introduced in the previous chapter. Eye-gaze response behaviour is analysed with quantitative data and fully illustrated with gaze plots of fixations and saccades, temporal plots of the x and y coordinates of gaze position (XY-T), and cumulative gaze activity heatmaps. Consideration is given to the reason why, at any instant, gaze might move to a new position and the influence this might have on RSVP design.

Keywords Eye-gaze behaviour • Variations in gaze behaviour • Static RSVP modes • Moving RSVP modes • Multiple entry/exit modes • Fixation plots • XY-T plots • Cumulative gaze heatmap plots

By means of the analytical tools and representations developed in the last chapter we now examine ten common RSVP modes selected from the three classes of static, moving and multiple entry/exit. We do so in order both to understand how gaze behaviour is influenced by the nature of each mode as well as to provide guidelines to an interaction designer who wishes to exploit these and other modes in an application.

5.1 Static Presentation Modes

First, we consider the difference in gaze behaviour between Slide-show mode RSVP, in which the user gets a single brief view of each image in a set and must recognise a specific image, and Tile mode, where all the images in the set are visible simultaneously. An interesting third example, Mixed (2×2) mode, combines aspects of both Slide-show and Tile mode. The gaze data used is that from the study by Cooper et al. (2006).

5.1.1 Slide-Show

It is immediately clear when analysing gaze behaviour patterns associated with Slide-show mode that the gaze point remains essentially fixed at the centre of the display area for the entire duration of the presentation. This is in direct contrast to the gaze behaviour that would be expected if the user were presented with each image individually for an extended period, as in Fig. 4.1, when a series of saccades and fixations around the image area would be expected. With only a short image exposure, we may presume that the user is reliant on rapid image “gist” recognition. Additionally, one might reasonably conjecture that centring on the image maximises the coverage of the higher resolution areas of the retina.

Based on gaze data from 10 users, Fig. 5.1 shows the cumulative heatmap gaze pattern for Slide-show mode, confirming the pattern of centralised steady gaze behaviour in this presentation mode. Not all individuals select the exact centre and Fig. 5.2 illustrates some of that variation.

Figure 5.2 (left) from participant #5 shows 5 fixations, all nearly closely co-located (average length 1,097 ms, longest 1,670 ms), whereas Fig. 5.2 (right) from participant #7 shows 12 rather more widely spread fixations in the same exposure time (average fixation length 418 ms, longest 850 ms). In the latter example the fixations are both

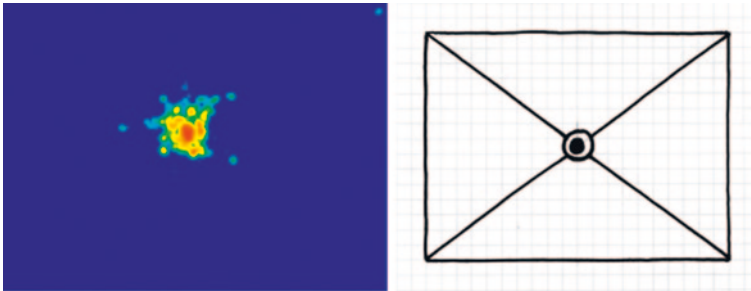


Fig. 5.1 Slide-show mode: cumulative heatmap (*left*), notation (*right*)



Fig. 5.2 Slide-show mode fixation pattern examples showing the natural variability between individuals

shorter and more scattered about the central point. This illustrates the difference in gaze stability between individuals—although there is no change in overall gaze strategy for this presentation mode from any tested individual.

5.1.2 Tile Mode

Tile mode (Fig. 5.3) presents all the images in the set, including any target images, simultaneously. In complete contrast to Slide-show mode, gaze behaviour adopts the visual search pattern characterised by short fixations interspersed by rapid saccades. Even though an explicit search task is being undertaken, this pattern of gaze activity is very similar to that observed when a user is simply viewing a single image of equivalent size.

The cumulative heatmap (Fig. 5.3, left) shows a generally even distribution of gaze over the image area. The few “hotspots” relate to apparently distinctive images (and target areas) that are visited more than others. Little is known about the mechanism by which the pattern of search is controlled, but it is almost never systematic or evenly distributed over the available search area.

Figure 5.4 (left) shows an illustrative example of gaze behaviour for Tile mode. During a presentation lasting 2.9 s, 11 fixations of average length 235 ms (longest 570 ms) were recorded. Figure 5.4 (right) shows the corresponding XY-T plot

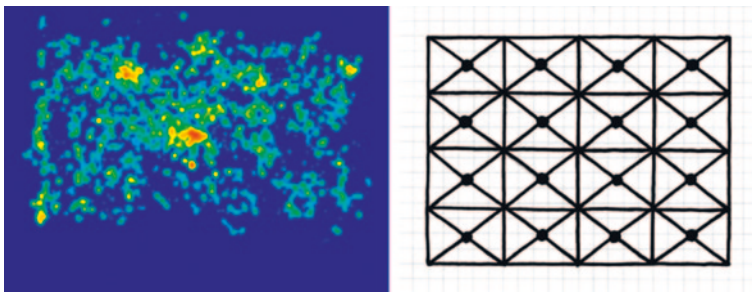


Fig. 5.3 Tile mode cumulative heatmap (*left*), notation (*right*), not to scale

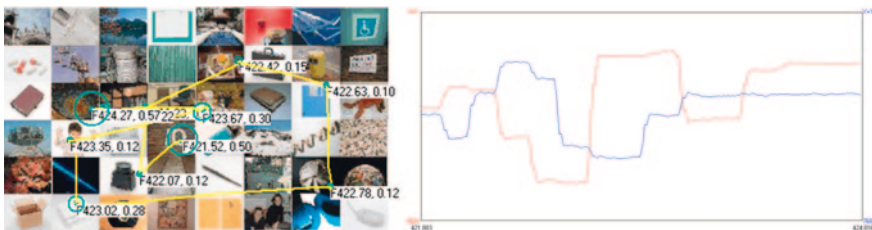


Fig. 5.4 Tile mode single exposure (2.9 s) (*left*), XY-T plot (*right*)



Fig. 5.5 Search to target (*left*), pre-attentive discovery (*right*)

representation. The pattern of saccades (steps) and fixations (plateaux) can clearly be seen. The target was not present in this example.

When a target image is present in Tile mode, it is sometimes detected after a period of visual search and sometimes immediately (“pre-attentively”). Figure 5.5 (left) shows a period of visual search prior to locating the target—the picture of a human eye in the rightmost column—with four distinct saccades and fixations before the target is acquired. The brief fixation just prior to a small “correction” onto the main target is relatively common after a long saccade and probably indicates a “falling short” by the ballistic saccadic estimation process (see Becker 1991, for a detailed description of the mechanism of saccades).

In contrast, the Tile mode trace shown in Fig. 5.5 (right) illustrates immediate acquisition of the target: the first complete saccade after the tiled image appears leads directly to a fixation on the target (the image of orange circle shapes). This example, we believe, indicates “pop-out” or pre-attentive recognition as discussed in Chap. 2. It occurs relatively frequently when a target is visually distinctive within the image set.

5.1.3 Mixed (2×2) Mode

Recall from Chap. 2 that Mixed mode was intended to provide longer viewing times for images, with a relatively minor, but compensating, reduction in size. Figure 5.6 (left) shows the cumulative heatmap for a range of examples of Mixed 2×2 mode, combined over the three different overall presentation rates (3.57, 2.50 and 1.92 mixed quad images per second) used by Cooper et al. (2006).

Figure 5.7 shows two illustrative examples of higher and lower presentation rates in 2×2 Mixed mode. With a longer overall presentation time (lower presentation rate) the user’s gaze typically moves in a non-regular pattern of saccades and fixations spread around the central region of the display area, Fig. 5.7 (left). As the presentation time decreases (rate correspondingly increases), Fig. 5.7 (right), the fixations become increasingly more centralised.

Overall this trend is consistent with an observed decrease in gaze travel speed with increase in presentation rate: 739 pix/s for a rate of 1.92 quad presentations/

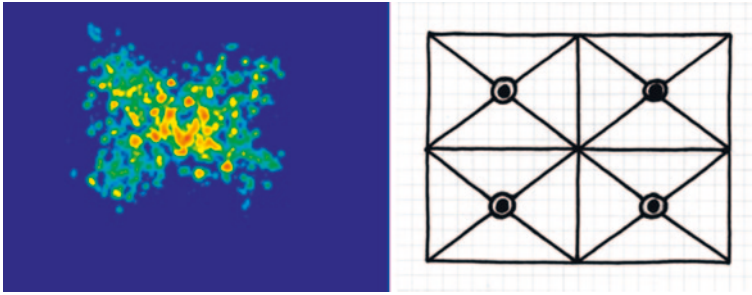


Fig. 5.6 Mixed 2×2 mode cumulative heatmap (left), notation (right)



Fig. 5.7 Slower presentation rate (left), higher presentation rate (right)

second (i.e. the 12 quad images in 6.24 s), as opposed to a gaze travel of 430 pix/s for a rate of 3.57 quad presentations/second (i.e. the 12 quad images in 3.36 s). It is not established whether very high presentation rates, comparable to that used in Slide-show mode, would give rise to fully centralised gaze.

5.1.4 Summary

Slide-show mode RSVP invariably evoked a steady gaze response. This response is similar to that observed for text RSVP and gives rise to low values of gaze travel; it is generally well liked by users and effective as a method of presentation. At high presentation rates blinking becomes a potential problem, as the perceptual blindness caused can completely mask a target image appearance.

Tile mode gave rise to a distinctive visual search strategy of saccades and fixations, as observed when viewing a normal picture on a display. We noted that in some instances the target image was noticed immediately (pre-attentively) and sometimes as a result of active search.

Mixed (2×2) mode, when presented comparatively slowly showed the characteristics of visual search mode, but becomes increasingly centralised and “steady” as image presentation rate increased. This mode was both successful, giving more time to view the images than Slide-show at a reasonable size, and well liked with low gaze travel rates. Due to extended presentation times, the effects of blinking are reduced.

5.2 Moving Modes

We now consider RSVP modes in which images have a substantive moving component. As with the static modes, changes in design have profound effects on the gaze behaviour adopted by users and these effects are not always as might be expected. Three modes are considered here: Ring, Stream and Diagonal mode, with data taken from the study by Cooper et al. (2006) and some additional observations about Diagonal mode from a study conducted with the University of Pavia in 2009.

5.2.1 Ring Mode

Recall that Ring mode provides a central area in which each image in the set appears, and is captured, as in Slide-show mode. Starting from the entry point, each image then moves in a spiral path leaving the display continuously at the top edge. The cumulative heatmap for this mode, Fig. 5.8 (left), indicates that the overwhelming majority of gaze activity is centralised and is of the steady gaze type. There are only very occasional excursions along the ring part, either following a very distinctive target or the last image as it leaves the screen.¹ Figure 5.9 shows an illustrative example; the average fixation length is 869 ms.

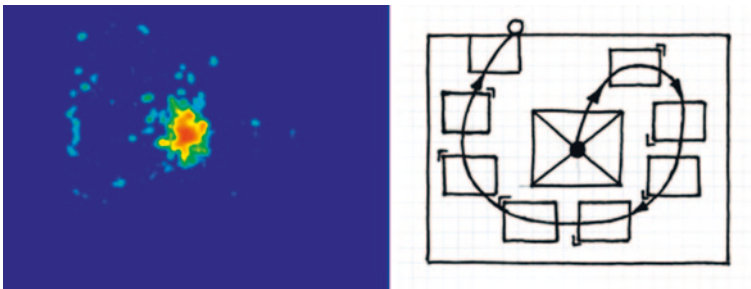
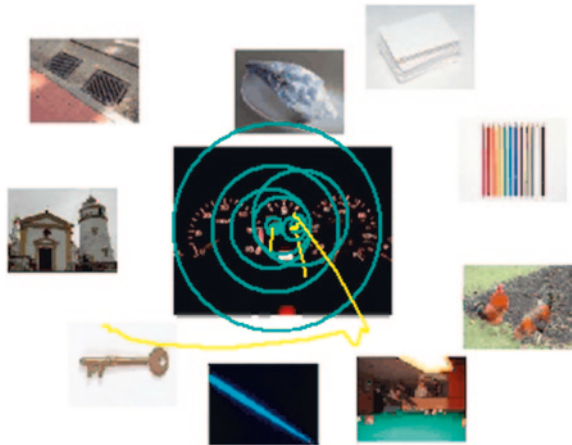


Fig. 5.8 Ring mode cumulative heatmap (*left*), notation (*right*)

¹ Cooper et al. (2006) report that at least one user adopted a nystagmic gaze behaviour in this Ring mode—coupled to a dislike of the mode!

Fig. 5.9 Ring mode, illustrative example from one presentation



The clear implication of this analysis is that the moving component of this Ring mode design plays no effective part in the RSVP process and may be considered essentially decorative.

5.2.2 Stream Mode

Stream mode differs from Ring mode in that there is no capture point at any location. Instead the images are constantly moving, appearing continuously at the right and moving along a curved path to the top, where they disappear continuously. Figure 5.10 (left) shows a cumulative heatmap for this mode. There is a concentration of gaze activity along the portion of the image path where the images are at their largest. The central hotspot represents the gaze position at the start of each sequence and has no other significance.

More detailed inspection of the gaze path reveals that eye movement is nystagmatic. Figure 5.11 (left) shows an illustrative example. Here, instead of fixations,

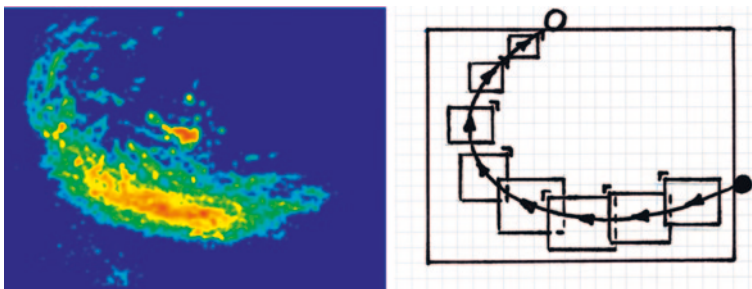


Fig. 5.10 Stream mode cumulative heatmap (left), notation (right)

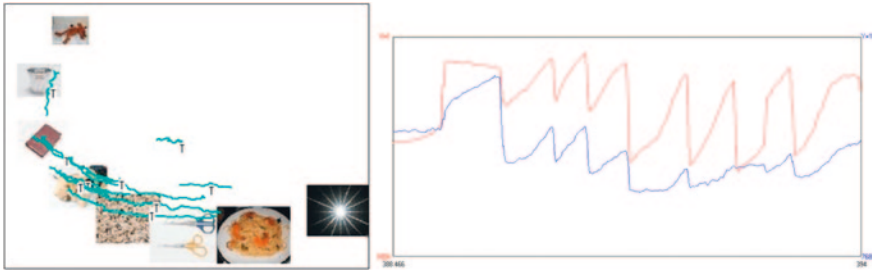


Fig. 5.11 Nystagmatic behaviour for one Stream mode sequence (*left*), XY-T plot (*right*)

a repetitive series of gaze path tracking events are observed (indicated with a “T”). Blue segments of the track indicate periods when successive gaze points are considered to comprise a single event. The usual yellow trace, representing the returning saccades, has been omitted for clarity. Figure 5.11 (right) shows the XY-T plot for the same sequence as Fig. 5.11 (left). The typical nystagmatic or sawtooth motion—ramp followed by rapid return—is characteristic of the image path tracking and saccadic returns just described. This gaze strategy—dominated by continuous gaze movements—is strongly linked to users’ reported dislike of this Stream mode.

5.2.3 Diagonal Mode

In Cooper et al. (2006) Diagonal mode example images continuously appear at the top left corner and rapidly move diagonally across the display, leaving at the bottom right hand corner in a capture frame. This Diagonal mode design has many similarities to the Stream mode just considered; but is gaze behaviour equivalent?

The cumulative heatmap from 10 participants (Fig. 5.12, left) indicates two distinct types of response by individual users. Two separate areas of activity are

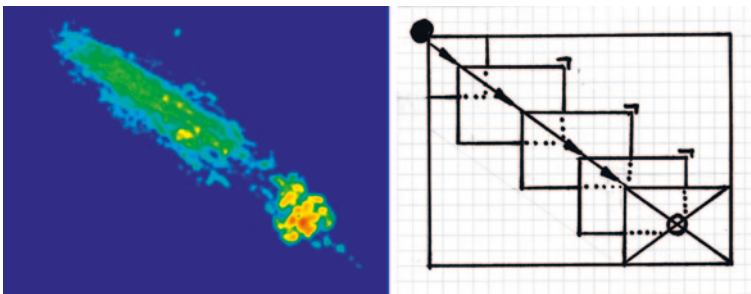


Fig. 5.12 Diagonal mode cumulative heatmap (10 participants) (*left*), notation (*right*)

apparent: first an elongated area diagonally across the screen, and second a group of locations near the exit point. In some cases, users will track the path of the moving image sequence nystagmatically in a manner very similar to the Stream mode response, but in about half of the cases gaze stabilises to a single steady gaze point near the bottom right capture frame.

Figure 5.13 (left) shows the response of one user: a series of closely spaced fixations near a single point. The associated XY-T plot confirms that this user moves from the (central) gaze point at the start of the sequence, follows the diagonal stream downwards and then remains steady near the bottom right corner (where the images are captured) until the end of the sequence. Figure 5.14 (left) shows the track for a single presentation sequence for a different user, showing the alternative nystagmatic strategy. This is confirmed by the XY-T plot (right).

Cooper et al. (2006) give no indication as to why users react differentially in this way in Diagonal mode, why it should be different from Stream mode, and whether there is any particular advantage gained by unconsciously adopting one or other of the gaze strategies. However, there are strong indications that users expressed greater preference for the Diagonal mode when they adopted the stable gaze strategy illustrated in Fig. 5.13.

In a separate study,² each user was presented with several instances of the Diagonal mode in succession (although with different paces and speeds) so that

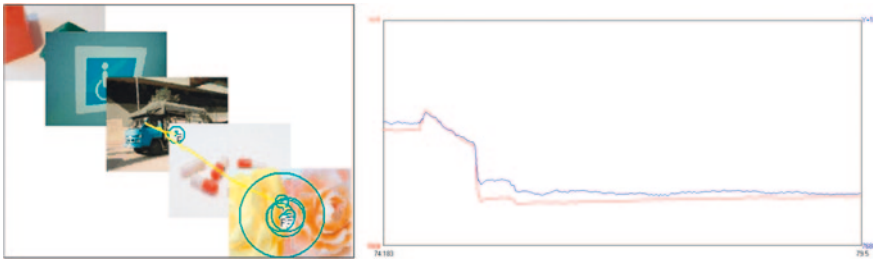


Fig. 5.13 Participant #10 with steady gaze (*left*), XY-T plot (*right*)

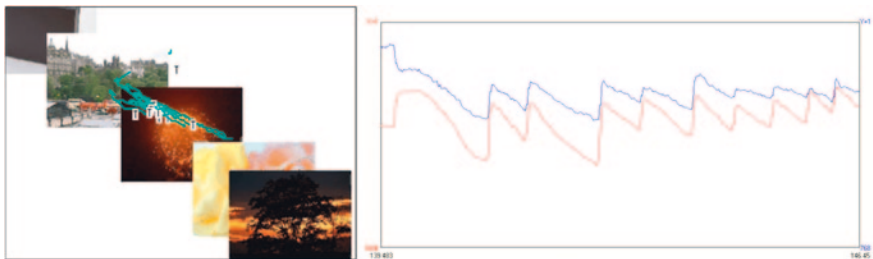


Fig. 5.14 Participant #11 showing nystagmus movement (*left*), XY-T plot (*right*)

² Conducted in collaboration with the University of Pavia.

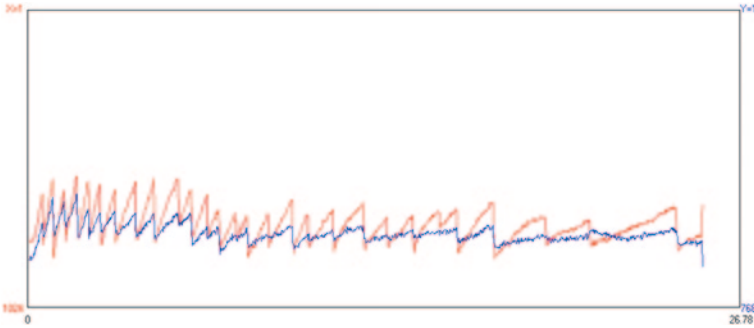


Fig. 5.15 Change from nystagmus towards stable gaze behaviour observed in diagonal mode

any change of strategy by an individual could easily be noted over the course of those presentations. Some users immediately demonstrated the steady gaze strategy (see Fig. 5.13) and some the nystagmus strategy (Fig. 5.14).

Much more clearly in the Pavia study than in Cooper et al's, there was a strong tendency for users to shift away from nystagmus tracking towards the steady gaze strategy over the course of the five individual presentations made to them. Once the steady gaze strategy had been adopted, there was no tendency to revert to the nystagmus strategy. Figure 5.15 shows a particular instance where the tracking was observed to slow markedly within a single sequence presentation, illustrative (but not typical) of the overall gaze strategy change shown by the majority of participants.

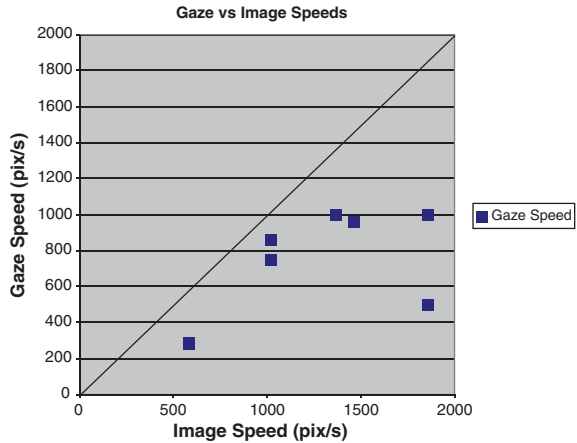
5.2.3.1 Effect of Image Speed and Limits to Gaze Tracking Behaviour

Figure 5.16 plots, for eight sample participants using Diagonal mode, the measured gaze speed across the screen against the actual speed of the image stream presented on the display. These are both expressed in pixels/second, and relate only to the X (horizontal) component of the gaze traces. The Y (vertical) component, however, demonstrates equivalent behaviour. Gaze speeds were manually estimated from the slope of the gaze traces.³

From this plot it generally appears that gaze speed nearly, but not quite, matches the image speed at lower presentation speeds, but does not exceed a gaze speed of approximately 1,000 pix/s. There appears to be a maximum angular speed for controlled gaze tracking, which will not be consistently

³ We also performed a linear regression on several instances of gaze track data, but these computed values were not sufficiently different from those obtained by manual estimation to cause any concern.

Fig. 5.16 Gaze speed vs. image speed



exceeded regardless of the speed of the image stream. While saccadic eye movements can demonstrate a much higher angular speed, they are ballistic and not controlled.

5.2.4 Comment

Ring mode, despite having a moving image stream, caused a steady gaze response similar to Slide-show mode. We conjectured, therefore, that the moving “ring” part was unlikely to play any significant role in the user’s decision making process and should therefore be considered as an ornamental rather than functional aspect of this mode.

Stream mode, comprised only of moving images with no capture frame, evoked rapid and continuous eye movements (nystagmus), which were neither effective nor liked by the users.

Similarly, Diagonal mode had a high movement component, but did not always give rise to nystagmus—users naturally tending to either a steady gaze or a nystagmatic strategy. We also noted that users naturally tended to transition to a steady gaze strategy if they initially adopted the nystagmatic one. This tendency is apparently independent of the presence of a capture frame.

There will almost always be one particular concern in the mind of the interaction designer, and that is an application’s visual appeal. While that cannot be quantified, it will nevertheless exert a strong influence on a design. Notwithstanding this concern with visual appeal, some clear advice can be drawn from the analyses just carried out. First, it appears that users favour a mode that allows them, for much of the time, to treat part of the presentation—that associated with a capture frame—as a Slide-show presentation. Useful gaze departures (i.e., other than nystagmatic) from such a location may be associated with context assessment or recognition confirmation, either or both of which may be beneficial to a specific application.

Similar conclusions may be drawn, and applied with caution, to moving modes other than diagonal, ring and stream, in which relatively small changes in a design can lead to radical shifts in gaze behaviour. In turn this will affect a user's performance and liking for a design. In this respect gaze recording and analysis represents a useful tool to support RSVP design.

5.3 Multiple Entry/Exit Modes

Perhaps the most interesting and thought provoking experimental results concerning the gaze behaviour and other features of multiple entry/exit modes was reported by Corsato et al. (2008). They conducted an experiment with five RSVP modes: Volcano, Floating, Shot, Collage and Grid.⁴ The task given to subjects was a theme target recognition task: out of a total of 2000 images, as many of them relevant to a given theme (e.g., 'dog', 'cat', 'aeroplane') were to be identified out of a possible total of 40 within the image collection, though every user was unaware of the latter number. The recorded results included total gaze travel, the degree of fatigue reported by users and the identification score out of 40. These three measures are represented for the four modes also shown in Fig. 5.17. Subsequently, the gaze records were accessed to generate heat maps.

Some outcomes of the experiment immediately stand out. First, the target recognition score is noticeably higher for the Volcano and Floating modes than for the other modes. Second, these two modes are characterised by the lowest extent by far of gaze travel. Third, Volcano and Floating modes are associated with the least degree of fatigue reported by users. Fourth, heat maps show that, for Volcano and Floating modes, gaze is quite localised spatially.

At first sight Corsato et al's Volcano mode would appear to have features in common with Slide-show and Ring modes, as well as aspects of Diagonal mode. At the centre is a simple sequential presentation of the total image set, as in Ring mode, followed by the images leaving the screen along eight separate paths. The fact that there are eight paths from the centre, rather than a single path as in Diagonal mode, has the effect of considerably slowing the image movement and reducing the need for overlap. In fact, analysis of gaze behaviour for Volcano mode reveals that neither the Slide-show nor Diagonal modes on their own provide an adequate model for understanding gaze behaviour.

Floating mode resembles Volcano, except that the images are serially presented centrally at a much smaller size and grow in size as they move towards the edge of the display screen. Shot mode is distinct from Volcano and Floating modes. Images appear at a point and gradually increase in size along various paths down the display screen, leaving continuously at the bottom of the display area.

⁴ Corsato et al's Grid mode is a close equivalent to Cooper et al's Tile mode and is omitted from consideration here.

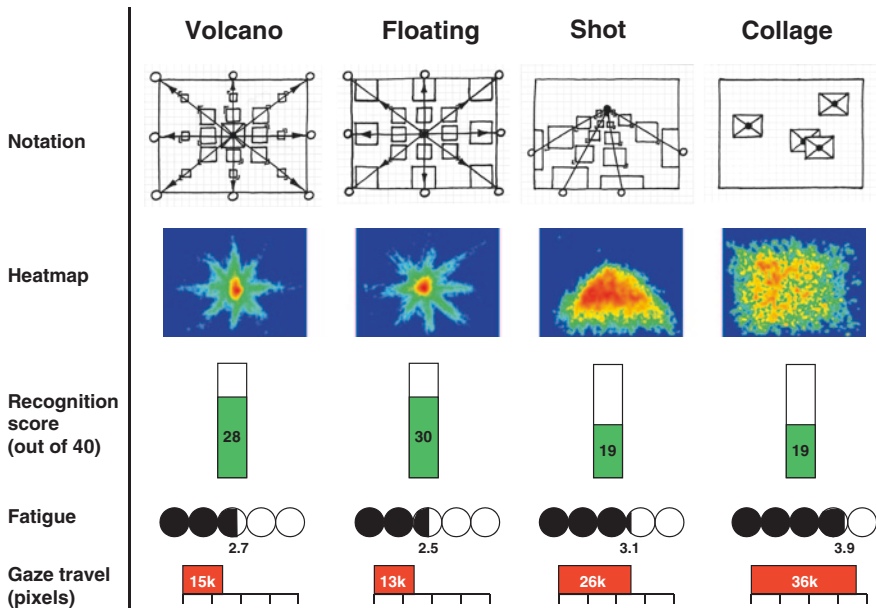


Fig. 5.17 Comparison of multiple entry/exit modes (data from Corsato et al. 2008)

Later we will investigate whether these characteristics of Shot mode gave rise to substantive differences in gaze behaviour.

5.3.1 Target Pursuit in Volcano, Floating and Shot Modes

Significantly, the gaze response of users to potential targets in these three multiple entry/exit modes is characterised by visual pursuit (refer back to Sect. 4.2.3) rather than the repetitive nystagmus pattern previously identified in Stream and Diagonal modes. In the three modes discussed in this section—Volcano, Floating and Shot—target appearance is strongly associated with individual visual pursuit events and these frequently result in target identification by user response. User responses are generally associated with correctly identified targets, but sometimes with a similar, but non-target image.

Figure 5.18 shows an example target pursuit event in Floating mode, in this case following the “dog” target image from the centre of the display to the right hand edge. The example is drawn from Floating mode, but gaze tracking follows a similar pattern in both Volcano and Shot modes. There is no immediate gaze response to the appearance of the image. Instead gaze can be seen to saccade (yellow) to the image as it moves along its designated path. Pursuit begins when visual tracking starts and ends with a distinct saccade. Where a target is identified

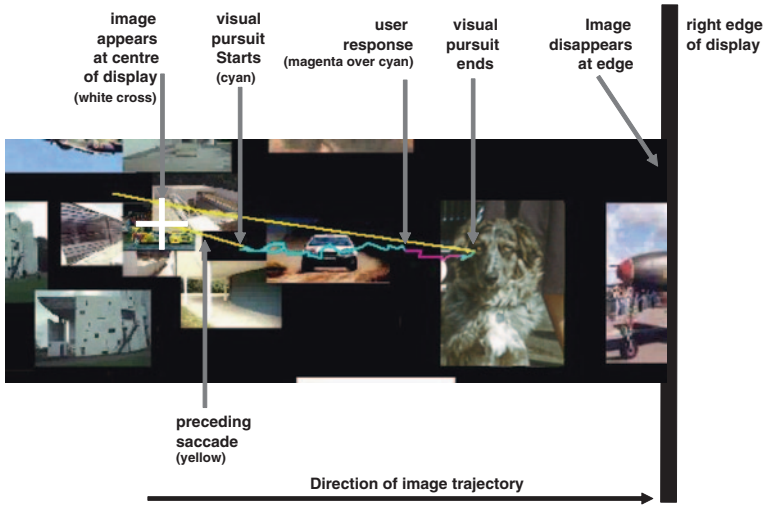


Fig. 5.18 An example target tracking visual pursuit in the floating mode (only part of the display is shown)

the user will indicate this by pressing a spacebar key (the user response). This is represented by a magenta coloured section of the overall cyan pursuit event.⁵ The user response usually occurs before the end of gaze pursuit, but is sometimes delayed until after it is finished.

For the three RSVP modes Fig. 5.19 shows diagrammatically the average timings and relative changes in image size as the images move along their defined paths. The timings were extracted by individual inspection of a video recording of the presentation overlaid with the gaze data representation.

We next look at gaze behaviour for these three selected multiple entry/exit modes in individual detail.

5.3.2 *Volcano Mode*

A cumulative heatmap representing five complete user trials for Volcano mode is shown in Fig. 5.20. The preponderance of steady gaze activity at the central point of appearance of the image stream is clearly shown, as are the radial pursuit events along the eight directions used.

Some idea of relative times can be drawn from measurements of 54 instances of volcano use. On average (see Fig. 5.19) visual pursuit began 900 ms after the

⁵ Only the start time of the user key press response is recorded in the experiment data, so the length of the magenta section is arbitrary and not significant here.

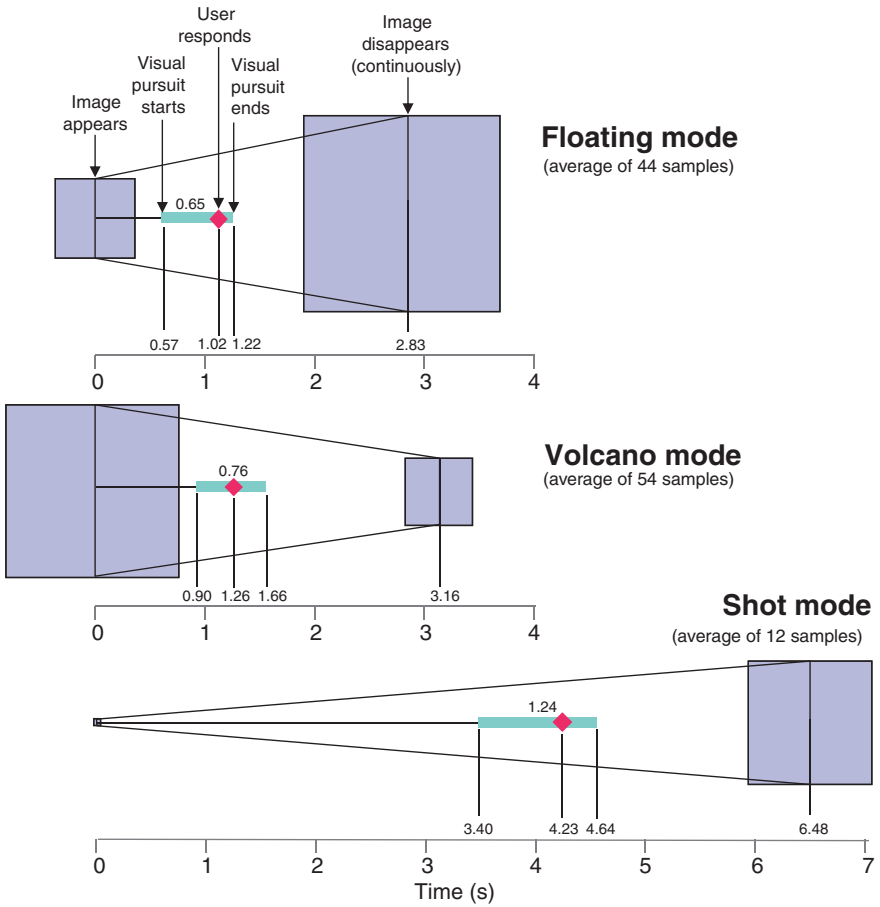


Fig. 5.19 Summary of pursuit timing and image size data

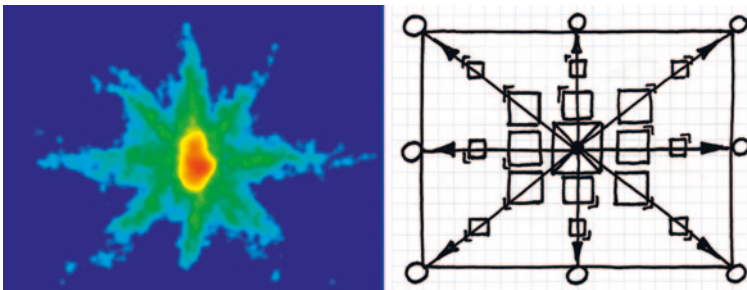


Fig. 5.20 Volcano mode cumulative heatmap (left, 5 trials), notation (right)

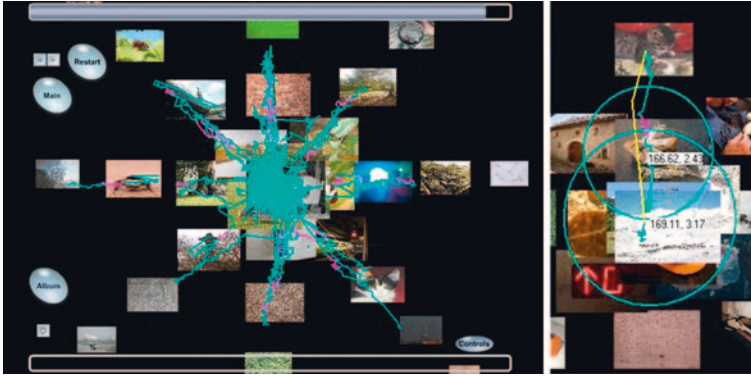


Fig. 5.21 Volcano mode, all tracks for one user (*left*), single pursuit, detail (*right*)

appearance of an image, and lasted 760 ms, though user response occurred 360 ms into the visual pursuit. The majority of user responses were made while pursuit was underway. On average, users continued to follow images for 400 ms after their response. There is, of course, a danger that closely spaced target examples may be missed if the user's attention is taken up following an existing target across the display when a second appears (there is at least one example of this occurring in the analysed data set). Figure 5.21 shows the tracking events for one user.

The variability of user behaviour is considerable. It is impossible to say, of course, whether a user is responding to a target image identified during steady gaze or during pursuit. In 3 cases out of the 54 the user responded before visual pursuit began. Sometimes users did not respond to a target and no gaze pursuit was recorded. At other times non-targets were followed but no response made: it is not clear if there was doubt in the user's mind or whether the image being followed was merely interesting.

5.3.3 Refining Volcano Mode

One of the clear findings from this study is that gaze behaviour for Volcano mode is not equivalent to Slide-show or Ring mode. Pursuit outside the central steady gaze region is the norm. This might be due to the relatively small size of the central image compared to Slide-show and Ring modes. It is a completely open question whether opting for a larger central image size would significantly alter the gaze process or target recognition rate and whether, if this were done, the moving radial image paths would then be used.

Figure 5.22 shows a possible conjectural redesign. The image size on appearance is made much larger, and the corresponding image sizes along the radial trajectories are larger also. Figure 5.19 indicates that the effective part of the display (i.e. where

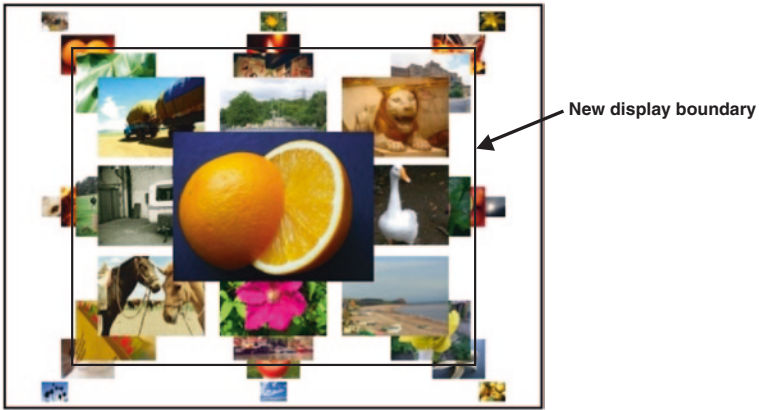


Fig. 5.22 Volcano mode, potential redesign (courtesy of T. Brinded)

recognition events have already occurred) is still well inside this new boundary, which would then be expanded to the edges of the display. Note, however, that the existing design is already quite effective from the point of view of target recognition (70 %, as reported by Corsato et al. 2008), and has a low fatigue rating (Fig. 5.17).

5.3.4 Floating Mode

Figure 5.23 (left) shows the cumulative heatmap for gaze activity for five trials of Floating mode, similar to that shown for Volcano mode (Fig. 5.20). The plots are apparently functionally similar—gaze is steady and concentrated at the centre where images appear (albeit smaller in this mode) and tracks outwards for potential targets (visual pursuit). It is interesting to note that the design changes between Volcano and Floating made little difference to the timings produced by the

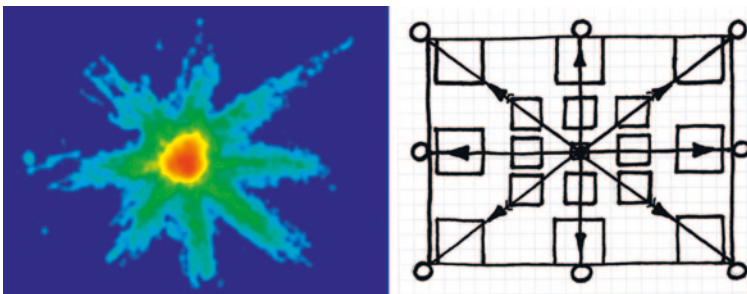


Fig. 5.23 Floating mode cumulative heatmap (5 trials) (left), notation (right)

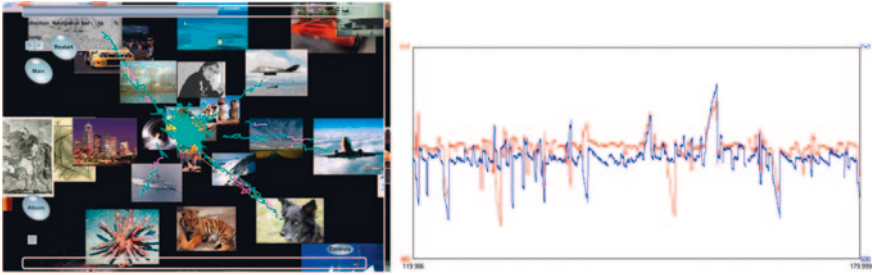


Fig. 5.24 Floating mode: selection of tracks from one user (60 s duration) (*left*), XY-T (*right*)

detailed analysis (summarised in Fig. 5.19). Gaze acquisition time is reduced, consistent with the reduced (and therefore less obscuring) size of the central image.

Figure 5.24 (left) shows a selection of tracks extracted from a 60 s segment taken from the complete 220 s sequence for one user. It may be seen that most of these pursuit tracks are directly associated with target recognition response—the magenta sections. Figure 5.24 (right) shows the XY-T plot for the 60 s extract. The larger visual pursuit movements are clearly visible, but the plot also indicates smaller tracking movements closer to the display centre.

5.3.5 Shot Mode

Corsato et al's Shot mode operates differently from Volcano and Floating modes. Images appear centrally and near the top of the display and move towards the bottom while getting larger. The cumulative heatmap (Fig. 5.25, left) for this mode shows that there is no gaze activity until the images have become a useable size.

Average pursuit start time for targets in this mode (Fig. 5.19) is 3.4 s and pursuit starts approximately halfway down the display (Fig. 5.25, left). Pursuit

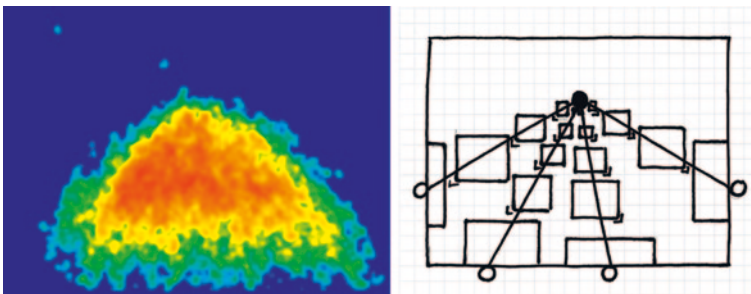


Fig. 5.25 Shot mode cumulative heatmap (5 trials) (*left*), notation (*right*)



Fig. 5.26 Shot mode, all tracks for one user

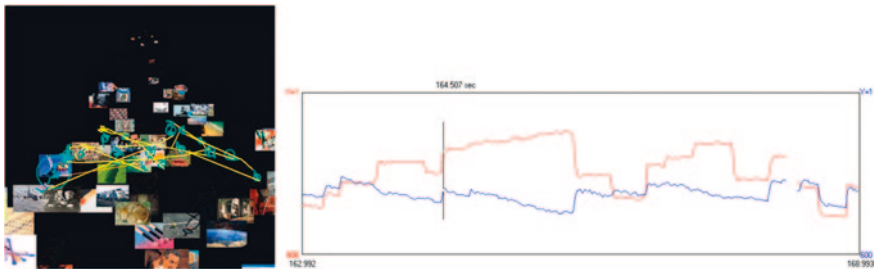


Fig. 5.27 Shot mode visual search behaviour detail (*left*), corresponding XY-T plot (*right*)

duration is also longer, on average, at 1.24 s. As with Volcano and Floating modes, there is clear visual pursuit behaviour along target image paths (Fig. 5.26), but in this mode there is also clear evidence of generally “horizontal” or side to side visual search between the pursuit events, rather than the steady gaze behaviour of Volcano and Floating modes.

Figure 5.27 (left) shows a six second detail of this horizontal side to side visual search behaviour. The user scans to and fro across the image area on the display, but generally stays within a relatively narrow horizontal band. Saccades range across the display width. Fixations are less stable than for a static image and tend to be somewhat elongated along the direction of the moving images. This six-second trace includes a single target pursuit event target identification (far left, large circle). Figure 5.27 (right) shows the XY-T plot for these 6 s. The single target pursuit event starts at 164.5 s (vertical line).

Each user appears to select a different image size range to search in this way, so that the central area of the cumulative heatmap appears spread vertically.

5.3.5.1 Refining Shot Mode

The analysis in the previous section indicates that half the display area in Shot mode is unused, and that the user is obliged to search the region of the display where images are at a minimum size for recognition. Altering the design so that it is reversed, with images appearing at their largest size at the bottom edge and tracking towards a point at the top of the display, would consistently provide a larger image for visual search while hardly compromising the user’s ability to pursue the images into the display. A separate study (Mardell et al. 2009) indicates that gaze naturally searches close to the edge of continuous entry (apparently regardless of top, bottom, left or right) and follows features of interest into the display area. Whether these changes would improve recognition or user preference remains an open question.

5.3.6 Collage Mode

In Corsato et al. (2008) Collage mode⁶ (illustrated previously in Fig. 3.12) images appear rapidly at random locations on the display and may be covered, completely or partially, at any time by a new image. The average time during which any given image was visible, either in its entirety, or partially obscured (see Sect. 3.3.3) was 6 s. However, the period during which images could be recognised was very variable, the shortest time for just over 300 ms, the longest for some 40 s.

The cumulative heatmap for this Collage mode is shown in Fig. 5.28. At first sight it might seem that the gaze pattern follows a conventional saccade and fixation visual search, apparently similar to that for Tile mode. However, in the more

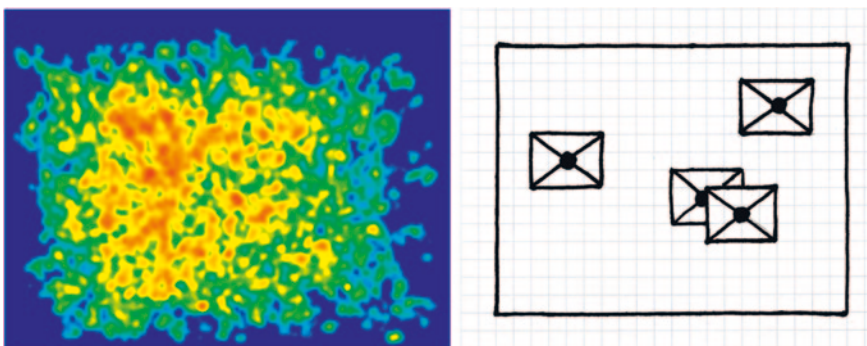


Fig. 5.28 Collage mode cumulative heatmap (5 trials), notation (*right*)

⁶ This Collage mode RSVP differed from the design by Wittenburg et al. (1998, 2000, illustrated in Fig. 1.3) in that a very large number of images are visible on the display at any one time.

detailed analysis of 36 target appearances (from a single user), the user responded correctly on just 12 occasions. In all but three of these, target recognition could be interpreted as pre-attentive in that the targets were fixated on and responded to within one second. The average response time from gaze acquisition to key press was 702 ms.

The remaining recognition instances relied on a normal visual search pattern, which varied considerably in the time taken—the longest delay between appearance and reported recognition being 31.5 s! This would imply that pre-attentive recognition is very important to success in this example of Collage mode, and search less effective. Interestingly, although there is clear evidence from the detailed analysis that while the user fixated on at least 19 of the available targets, only 12 were positively identified. There was no instance of an identification response made without the user first fixating on the target image.

Collage mode is different from Tile mode in that pre-attentive recognition appears to be much more significant, yet there appears to be a conflict between searching and the potential interruptions posed by the continuous stream of new images. New images appear too quickly to be fixated on individually in turn such that the balance between search and immediate recognition must be maintained. Generally this example of Collage mode was neither successful in terms of effective recognition nor, with its extensive gaze travel, particularly popular with users.

5.3.7 Multiple Entry/Exit Modes, A Summary

Both Volcano and Floating modes showed a combination of steady gaze with interruptions due to visual pursuit events tracking potential targets across the display.

Shot mode also exhibited the visual pursuit of potential targets, but gaze showed a distinct tendency to a visual search strategy across the spread of images as they became large enough to identify. Shot mode was less effective and less popular than Volcano and Floating, showing a greater extent of gaze travel.

The gaze behaviour invoked by Collage mode, while superficially similar to the visual search strategy adopted in Tile mode, was both ineffective and unpopular—careful analysis showed a strong reliance on pre-attentive recognition, which was in turn often masked by the rapid appearance of distracting new images.

The most favoured and effective modes (Floating and Volcano) were also found to be associated with shorter gaze travel paths compared with the Shot and Collage modes. With regard to the number of correct images selected, Floating and Volcano scored highest, with Shot much lower. User preference was noticeably higher, and fatigue much lower, for Floating and Volcano modes compared with Shot and Collage modes. Thus, for all Corsato et al's measures of performance in these multiple entry/exit modes, Floating and Volcano were found to be superior to Shot and Collage.

5.4 Summary

This chapter has concentrated on the gaze responses of actual RSVP users. Previously, we identified four gaze strategies: visual search, steady gaze, nystagmus and visual pursuit, which, taken together, characterise our users' natural gaze responses to 10 very different RSVP designs.

In all, gaze analysis can give the RSVP designer significant insights into how a new design will be accepted by users. The designer can use eye gaze technology to refine designs to avoid pitfalls arising from perceptual artefacts.

References

- Becker, W. (1991). Saccades. In R. H. S. Carpenter (Ed.), *Eye movements* (Vol. 8, pp. 95–137). Vision and visual dysfunction Boca Raton: CRC Press.
- Cooper, K., de Bruijn, O., Spence, R., & Witkowski, M. (2006). A Comparison of static and moving presentation modes for image collections (pp. 381–388). *Proceedings of Advanced Visual Interfaces (AVI-2006)*.
- Corsato, S., Mosconi, M., & Porta, M. (2008, May). An eye tracking approach to image search activities using rsvp display techniques, (pp. 416–420), *ACM, Proceedings of Workshop on Advanced Visual Interfaces, Naples*.
- Mardell, J., Witkowski, M., & Spence, R. (2009). Detecting Search and rescue targets in moving aerial images using eye-gaze (pp. 67–70). *Proceedings of 5th International Conference on Communication by Gaze Interaction (COGAIN-09)*.
- Wittenburg, K., Ali-Ahmad, W., LaLiberte, D., & Lanning, T. (1998). Rapid-fire image previews for information navigation (pp. 76–82). *ACM, Proceeding Conference on AVI*.
- Wittenburg, K., Chiyoda, C., Heinrichs, M., & Lanning, T. (2000, January 26–28). Browsing through rapid-fire imaging: requirements and industry initiatives. In *Proceedings of Electronic Imaging '2000: Internet Imaging* (pp. 48–56). San Jose, CA, USA.